```
                GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    August 28, 2002, 07:07:20 ; Search time 1593.33 Seconds
                                        (without alignments)
                                        13138.381 Million cell updates/sec

Title:          US-08-711-417C-165
Perfect score:  1551
Sequence:       1 ATGGATGCTGACGAGGTCA.........ACCGCTTCCACATGAGCTAA 1551

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:    27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  EST:*
            1:  em_estba:*
            2:  em_esthum:*
            3:  em_estin:*
            4:  em_estmu:*
            5:  em_estov:*
            6:  em_estpl:*
            7:  em_estro:*
            8:  em_htc:*
            9:  gb_est1:*
            10: gb_est2:*
            11: gb_htc:*
            12: gb_gss:*
            13: em_gss_hum:*
            14: em_gss_inv:*
            15: em_gss_pln:*
            16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	452.4	29.2	994	9	AI325349	AI325349 mi06e03.y
2	424.2	27.4	994	9	BF794111	BF794111 60225675
3	409.2	26.4	948	9	AL561534	AL561534 AL561534
4	383	24.7	442	9	AW503638	AW503638 UI-HF-BN0
c 5	372.8	24.0	509	9	AA027561	AA027561 mi06e03.r
6	362	23.3	731	9	BB207438	BB207438 BB207438
7	353.6	22.8	669	9	BB629847	BB629847 BB629847
c 8	348.8	22.5	936	9	AA866707	AA866707 vx85a10.r
c 9	325.4	21.0	339	9	AA808413	AA808413 oa43h06.s
c 10	271.4	17.5	369	9	AA814418	AA814418 nz12d08.s
c 11	251.6	16.2	293	9	AW976805	AW976805 EST388914
12	239.2	15.4	1093	10	BM458447	BM458447 AGENCOURT
c 13	237.4	15.3	704	10	BF731127	BF731127 mab81el1.
14	231.8	14.9	571	10	BG089790	BG089790 mab81el1.
15	229.8	14.8	451	9	AI323339	AI323339 mi06e03.x
16	221.4	14.3	686	9	AJ396733	AJ396733 AJ396733
17	198.8	12.8	648	10	BF471977	BF471977 UI-M-BH3-
c 18	191.4	12.3	577	9	AA863970	AA863970 vx87d08.r
19	179.8	11.6	708	9	BB635722	BB635722
20	176.4	11.4	404	9	AA223756	AA223756 zr10c01.r
21	174.6	11.3	300	9	BB489077	BB489077
22	172	11.1	381	9	AW502759	AW502759 UI-HF-BR0
23	166	10.7	304	9	BB235180	BB235180 BB235180
24	164	10.6	566	10	BI680684	BI680684 459016 MA
25	162.6	10.5	529	10	BM362676	BM362676 BS320470
26	159.8	10.3	665	9	AJ399435	AJ399435 AJ399435
27	158.6	10.2	674	9	BB634826	BB634826
28	156.8	10.1	291	9	BB230132	BB230132
29	149.8	9.7	434	9	AJ393522	AJ393522
30	149.6	9.6	863	11	AK004152	AK004152 Mus muscu
31	148.8	9.6	358	9	AW405039	AW405039 UI-HF-BL0
32	148.2	9.6	515	10	BM148203	BM148203 TCAAP1Q95
33	147.4	9.5	316	9	BB491059	BB491059
34	147.4	9.5	497	9	AJ397064	AJ397064
35	143.2	9.2	440	10	BG094579	BG094579 ut72d06.x
36	142.2	9.2	471	10	BG145166	BG145166 ut76e04.y
37	140.8	9.1	468	10	BE947123	BE947123 UI-M-BH3-
38	134.6	8.7	2055	11	AF116605	AF116605 Homo sapi
c 39	129	8.3	285	9	BB231819	BB231819
c 40	126.8	8.2	430	9	AI241390	AI241390 qh74b07.x
41	124.2	8.0	235	9	AI365526	AI365526 qz07g02.x
42	118.8	7.7	653	9	BB588267	BB588267
43	118.6	7.6	297	9	BB488934	BB488934
44	115	7.4	401	9	AI644775	AI644775 vb16d03.y
45	115	7.4	517	9	AA290536	AA290536 vb16d03.r

ALIGNMENTS

```
RESULT    1
AI325349/c
LOCUS       AI325349                557 bp    mRNA    linear   EST 23-DEC-1998
DEFINITION  mi06e03.y1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA
            clone IMAGE:459676 5' similar to gb:L03547 Mouse Ikaros DNA binding
            protein (MOUSE);, mRNA sequence.
ACCESSION   AI325349
VERSION     AI325349.1  GI:4059778
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 557)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R., and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:276564
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1..557
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
```

RESULT 1 (continued)

```
                  /db_xref="taxon:10090"
                  /clone="IMAGE:459676"
                  /clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
                  /sex="unknown"
                  /tissue_type="placenta"
                  /dev_stage="adult"
                  /lab_host="DH10B"
                  /note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
                  with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                  1st strand cDNA was primed with a Not I - oligo(dT) primer
                  [5'
                  TGTTACCAATCTGAAGTGGGAGGCGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTTT
                  T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                  (Pharmacia), digested with Not I and cloned into the Not I
                  and Eco RI sites of the modified pT7T3 vector. library
                  went through one round of normalization, and was
                  constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT       116 a    157 c    127 g    155 t
ORIGIN

Query Match             29.2%;
Best Local Similarity   89.5%;   Score 452.4;   DB 9;   Length 557;
                                 Pred. No. 7.9e-84;
Matches   486;  Conservative    0;   Mismatches  57;   Indels   0;   Gaps   0;

Qy      1 ATGGATGCTGACGAGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAGGAAAGCCCCCTGTA 60
          ||||||| || ||||||||||||||||||||
Db    543 ATGGATGTCGATGAGGTCAAGACATGTCCCAAGTTTCAGGAAAGGAGGCCCCCAGTC 484

Qy     61 AGCGATACTCCAGATGAGGGCGATGAGCCCATGCCCGAGGACCTCTCCACCACC 120
          || |||| |||||||||||||||| || ||||||||||||||| ||||| |||
Db    483 AGTGACACTCAGATGAAAGGGGATGAGCCCATGCCCATGCCCTGTCCCACTACC 424

Qy    121 TCGGGAGGACAGCAAAGCTCCAAGAGTGACAGATCGTCGGCCAGTAATGTTAAAGTAGAG 180
          || |||| |||||||||||||||||||||   ||| ||||||||||||||||||||||
Db    423 TCTGGAGCACAGCAGCAGAACTCCAAGAGTGATCGAGGCCATGGCCAGTAATGTTAAAGTAGAG 364

Qy    181 ACTCAGAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGAAGAATGTGCGGAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 ACTCAGAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGAAGAATGTGCAGAG 304

Qy    241 GATTTACGAATGCTTGATGCCTCGGGAGGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    303 GATTTACGAATGCTTGATGCCTCGGGAGGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 244

Qy    301 AGCTCGGCCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTAAAGTGTGAT 360
          |||||||| |||||||||||||||||||||||||||||||| |||||||||||||||||
Db    243 AGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTAAAGTGTGAT 184

Qy    361 ATCTGTGGGATCATTTGCATCGGGCCGCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACT 420
          || |||||||| || ||||||||||  |||||||||||||||||||
Db    183 ATCTGTGGGATCGTTTGCATCGGGCCCCAATGTCTCATGGTTCACAAAAGAAGTCATACT 124

Qy    421 GGAGAACGGCCCTTCCCAGTGCAATCAGTGCGGGGCCTCATTCACCCGAGAAAGGGCAACCTG 480
          || ||||||| ||||| |||||||||||||| ||||| |||
Db    123 GGTGAACGGCCCTTTCCAGTGCAACCAGTGTGGGGCGTCCTTCTTTACCCGAGAAAGGCAACCTC 64

Qy    481 CTCCCGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTAC 540
          || ||||||||||||  || || ||||||
Db     63 CTGCGGCACATCAAGCAAGCTGCACTCGGGGTGAGAGAGCCCTTCAAATGCCATCTTTGCAACTAT 4

Qy    541 GCC 543
          |||
Db      3 GCC 1
```

RESULT 2

```
BF794111
LOCUS       BF794111        994 bp    mRNA     linear   EST 12-JAN-2001
DEFINITION  602255675F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4338834 5',
            mRNA sequence.
ACCESSION   BF794111
VERSION     BF794111.1   GI:12099165

KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 994)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9948 row: f column: 19
            High quality sequence stop: 657.

FEATURES             Location/Qualifiers
     source          1..994
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4338834"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC library."

BASE COUNT       302 a    234 c    306 g    152 t
ORIGIN

Query Match             27.4%;
Best Local Similarity   80.6%;   Score 424.2;   DB 10;   Length 994;
                                 Pred. No. 6.4e-78;
Matches   570;  Conservative    0;   Mismatches 128;   Indels   9;   Gaps   6;

Qy      1 ATGGATGCTGACGAGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAGGAAAGCCCCCTGTA 60
          |||||||||||| ||||||||||||||||          ||||||||||||||||||||||
Db    192 ATGGATGCTGATGAGGTCAAGACATGTCCCAAGTTTCAGGGAAGGAAGGAAAGCCCCCCTGTA 251

Qy     61 AGCGATACTCCAGATGAGGGCGATGAGCCCATGCCCGAGGACCTCTCCACCACC 120
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    252 AGCGATATCCAGATGAGGGCGATGAGCCCATGCCCGAGGACCTCTCCACCACC 311

Qy    121 TCGGGAGGACAGCAAAGCTCCAAGAGTGACAGATCGTCGGCCAGTAATGTTAAAGTAGAG 180
          ||||||||||||||||||||||||||||||||| ||||| |||||||||||||||||||
Db    312 TCGGGAGGACAGCAAAGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAG 371

Qy    181 ACTCAGAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGAAGAATGTGCGGAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 ACTCAGAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGAAGAATGTGCGGAG 431

Qy    241 GATTTACGAATGCTTGATGCCTCGGGAGGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 GATTTACGAATGCTTGATGCCTCGGGAGGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 491

Qy    301 AGCTCGGCCTTTGTCGGGAGTTGGAGGCATTCGACTTCCT-AACGGAAAACTAAAGTGTGA 359
          |||||||| |||||||||||||||||||||||||||||||  ||||||||||||||||||
Db    492 AGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAAACGGAAAACTAAAGTGTGA 551

Qy    360 TATCTGTGGGATCATTTGCATCGGGC 419
          |||||||||||||||||||||||||
Db    552 TATCTGTGGGATCATTTGCATCGGG-CCAATGCTGCATCGGG 610

Qy    420 TGAGAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCCAGAAGGGCAACCT 479
          |||||||| |||||||||||||||||||||||| ||||||||
Db    611 TGGGAACGG-CCTTCAGTGCAATCAGTGCGGGG-CTCATTCAGTGCAATCAGTGCGGGGCA--CT 666
```

```
Qy  480  GCTCCGGCACATCAAGCTGCATTCCGGGGAGGAAGCCCTTCAAATGCCACCTCTGCAACTA  539
         ||||||||||||| |||| ||| |  ||  || |  |||| || ||||| ||| ||||||||
Db  667  GGTCCGGCACATCATGCTGGATCCGGGAGAAA---CCCTTAAATGTCACTTCTGAATAAG   723

Qy  540  CGCCTGCCCGCGGAGGAACGCCCTCACTGCACCTGGCCACCTCCGTTGGTAAACC        599
         | |  | ||  ||||||| || || | | ||||   ||| |||  ||| ||||||
Db  724  CATGTCGCCGAGGAACGCCCTACTGCCACCTGGCCACCTGGAGGAGCCACGACATAAGAGAGAAATA  783

Qy  600  TCACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGAGGAACATAA  659
         ||    |   ||   |  |||  |||||    |||  || ||   |  |||  ||||
Db  784  CAGGAATGGGAGAGACCTGGAGAGAAGAAGACCCGGGAACAAAACAGAAACCGCCACAA    843

Qy  660  AGAGCGCGTGCCACAACTACTTGGAAAGCCATGGGCCTTCCGGCCACAC  706
             ||| | ||  || |  ||| || ||  |||||| || ||||
Db  844  CGAGCCCCTGGCCCGCACACGGGGGAAGGCCGTGCCACGCCGGAGAGAC  890
```

RESULT 3
AL561534

```
LOCUS       AL561534              948 bp    mRNA      linear    EST 16-FEB-2001
DEFINITION  AL561534 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL011YO03 5
            prime, mRNA sequence.
ACCESSION   AL561534
VERSION     AL561534.1  GI:12909059
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 948)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..948
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DL011YO03"
                     /clone_lib="LTI_NFL010_BC2"
                     /sex="male"
                     /tissue_type="B cells from Burkitt lymphoma"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL:
                     http://fulllength.invitrogen.com"
```

```
BASE COUNT    251 a    212 c    269 g    215 t    1 others
ORIGIN
```

Query Match   26.4%;   Score 409.2;   DB 9;   Length 948;
Best Local Similarity 98.1%;   Pred. No. 8.2e-75;
Matches 414;   Conservative 0;   Mismatches 8;   Indels 0;   Gaps 0;

```
Qy    1  ATGGATGCTGACGAGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAGCCCCCCTGTA   60
         |||||||||||| ||||||||||||||||||   |||||||||||||||||||||||||
Db  155  ATGGATGCTGATGAGGTCAAGACATGTCCCAAGTTCCCAAGTTCAGGGAAGGAAAGCCCCCCTGTA  214

Qy   61  AGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCCCCGAGGACCTCTCCACCACC  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  215  AGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCCCCGAGGACCTCTCCACCACC  274

Qy  121  TCGGGAGGAGCAGCAAAGCTCCAAGAGTCGCAGAGAGTCGTGGCCAGTAATGTTAAAGTAGAG  180
         |||||||| ||||||||||||||||||| ||||| ||||||||||||||| ||||||||||||
```

```
Db  275  TCGGGAGGACAGCAAAGCTCCAAGAGTCACAGAGTCGTGGCCAGTAATGTGAAAGTAGAG  334

Qy  181  ACTCAGAGTGATGAAGAGAATGGCGTGCCTCTGAAATGATGGGAAGAATGTGCGGAG    240
         |||||||||||||||||||||||  ||||||  ||||||||||||||||||||||||
Db  335  ACTCAGAGTGATGAAGAGAATGGGCGTGCCTGCTGAAATGATGGGAAGAATGTGCGGAG  394

Qy  241  GATTACCAAATGCTTGATGCCTCCGGAGAGAAAATGAATGGCTCCCAGGGACCAAGGC   300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  395  GATTACCAAATGCTTGATGCCTCCGGAGAGAAAATGAATGGCTCCCAGGGACCAAGGC  454

Qy  301  AGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTAAAGTGTGAT  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  455  AGCTCGGCTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTAAAGTGTGAT  514

Qy  361  ATCTGTGGGGATCATTTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACT  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  515  ATCTGTGGGGATCATTTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACT  574

Qy  421  GG  422
         ||
Db  575  GG  576
```

RESULT 4
AW503638

```
LOCUS       AW503638              442 bp    mRNA      linear    EST 01-MAR-2000
DEFINITION  UI-HF-BN0-ald-e-10-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
            IMAGE:3079386 5', mRNA sequence.
ACCESSION   AW503638
VERSION     AW503638.1  GI:7140212
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 442)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..442
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3079386"
                     /clone_lib="NIH_MGC_50"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (3.5-4.4kb). Directionally cloned. Cells provided by
                     Louis M. Staudt, Ph.D. Library preparation by Maria de
                     Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "
```

```
BASE COUNT    87 a    165 c    138 g    52 t
ORIGIN
```

Query Match   24.7%;   Score 383;   DB 9;   Length 442;
Best Local Similarity 95.2%;   Pred. No. 1.8e-69;
Matches 418;   Conservative 0;   Mismatches 15;   Indels 6;   Gaps 2;

```
Qy   964 GGGCCGAGTCCCTGCCCGCTGGTCAGACGCCCCGGGGGGTTCCGAGTGGTCCCG 1023
         | |||||||||||||||||||||||  | |||||||||| ||||||||||||||
Db     4 GAGGCCGAGTCCCTGCCCGCTGGTGCACACGCCCCGGGCGGTTCCGAGTGGTCCCG 63

Qy  1024 GTCATCAGCCCGATGTACCAGCTGCACA---GGCGCTCGGAGGGCACCCGCGCTCCAAC 1080
         ||||||||||||||||||||||||||||    |||||||||||||||||||||||||||
Db    64 GTCATCAGCCCGATGTACCAGCTGCACAAGCTGCGCTCGGAGGGCACCCGCGCTCCAAC 123

Qy  1081 CACTCGGCCCAGGACAGCCGCGTGGAGTACCTGCGTCGTGCTCTCCAAGGCCAAGTTGGTG 1140
         |||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db   124 CACTCGGCCCAGGACAGCCGCGTGGAGAACCTGCGTCGTGCTCTCCAAGGCCAAGTTGGTG 183

Qy  1141 CCCTCGGAGGCGCGAGGGCTCCCCGAGGCCAAGACTGCCAAGACTCCACGGACACCGAGAGC 1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CCCTCGGAGGCGCGAGGGCTCCCCGAGGCCAAGACTGCCAAGACTCCACGGACACCGAGAGC 243

Qy  1201 AACAACGAGGAGGAGCCAGCCAGCCGTCTTATCACCTGACCAACCACACATCGCCCGACGGCCG 1260
         |||||  ||||||||||||||||||||| | ||| |||||||||||||| | ||||| | ||| 
Db   244 AACAAGGAGGAGGAGCCAGCCAGCCGTCTCATCTACCTGACCAACCACATCGCCCCGCACGCG 303

Qy  1261 C---AACGGCGTCGCGTCCGCTCAAGGAGGAGGAGCCACACCGCCCCGACCTGCTGCGCGGCCGCCTCC 1317
         |    |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 CGCAACGGGCTGTCGCGTCCGCTCAAGGAGGAGGAGCCACACCGCCCCGACCTGCTGCGCGGCCGCCTCC 363

Qy  1318 GAGAACTCGCAGGACGACGCGCTCCCGCGTGGTCAGCACCAGCGGGAGCAGATGAAGTGTAC 1377
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   364 GAGAACTCGCAGGACGACGCGCTCCCGCGTGGTCAGCACCAGCGGGAGCAGATGAATGTGTAC 423

Qy  1378 AAGTGCGAACACTGCCGGG 1396
         |||||||||||||||||||
Db   424 AAGTGCGAACACTGCCGGG 442
```

```
RESULT 5
AA027561/c
LOCUS       AA027561      509 bp    mRNA    linear    EST 21-JAN-1997
DEFINITION  m106e03.r1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA
            clone IMAGE:459676 5' similar to gb:L03547 Mouse Ikaros DNA binding
            protein (MOUSE); mRNA sequence.
ACCESSION   AA027561
VERSION     AA027561.1  GI:1493554
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 509)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:276564
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1..509
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:459876"
                     /clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
                     /sex="unknown"
                     /tissue_type="placenta"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACCAATCTGAAGTGGAGGAGCGGCGCCGGAAATTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT    106 a   143 c   114 g   146 t
ORIGIN
```

```
Query Match         24.0%;  Score 372.8; DB 9; Length 509;
Best Local Similarity   89.7%;  Pred. NO. 2.5e-67;
Matches 434;  Conservative 0;  Mismatches 47;  Indels 3;  Gaps 3;

Qy   16 GGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAGCCCCCTGTAAGCGATACTCCAGAT 75
         |||||||||||| |   ||| | ||||| |  ||||| | |   |  | ||||||||||
Db  481 GGTCAAGACATGTCCCAAGTTCAGAACAG-AAGCCCCACAGTCAGTGACACTCCAGAT 423

Qy   76 GAGGGCGATGAGCCCGATCCCGAGGACCTCTCCACCACCTCGGGAGGACAGCAA 135
         || ||  ||||||||||| || | |||| |||||| |||| |||||||||||||
Db  422 GAAGGGGATGAGCCCGATGCCTGTCCCTGAGGACCTGTCCACTACCTCTGGAGCACCAGCAG 363

Qy  136 AGCTCCAAGAGTGACAGAGTCGTGCCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAA 195
         | |||||||||||| |  |   ||||||||||||||||||||||||||||||||||||||
Db  362 AACTCCAAGAGTGATCGCAGCATGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAA 303

Qy  196 GAGAATGGGCGTGCCTGCTGGAAATGAATGGGAAGAAGAATGTGCGGAAGATTTACGAATGCTT 255
         ||||||||||||||||| || |||||||||||||||||||||||| |||| ||||||||||||
Db  302 GAGAATGGGCGTGCCTGTGAAATGAATGGGAAGAAGAATGTGCAGAGATTTACGAATGCTT 243

Qy  256 GATGCCTCGGGAGAGAAAATGAATGGCTCCCAGGGACCAAGGCAGCTCGGCTTTGTCG 315
         ||||||||||||||||||||||||||||  || ||||||||||||||||||||||||
Db  242 GATGCCTCGGGAGAGAAAATGAATGGCT-CCACAGGGACCAAGGCAGCTCGGCTTTGTCA 184

Qy  316 GGAGTTGGAGGCATTCGACTTCCTAACGACAAAACTAAAGTGTGATATCTGTGGGATCATT 375
         |||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||| ||
Db  183 GGAGTTGGAGGCATTCGACTATCCAACGACAAAACTAAAGTGTGATATCTGTGGGATCGTT 124

Qy  376 TGCATCTGGGCCCAATGTGCTCATGGTTCACAAAGAGAGCCACACTGGAGAACGGGCCCTTC 435
         |||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||| |||
Db  123 TGCATCTGGGCCCAATGTGCTCATGGTTCACAAAGAGAGCCATACTGGTGAACGGGCCTTTC 64

Qy  436 CAGTGCAATCAGTGCGGGGGCCTCATTCACCCGAAGAGGCAACCTGCCTCCGGCACATCAAG 495
         |||||||| |||||  ||||||||| ||||||||||||||||||| | ||||| |||||||
Db   63 CAGTGCAACCAGTGTGGGGGCCTCCTTCACCCGAAGAGGCAACCT-CTGCGGCACATCAAG 5

Qy  496 CTGC 499
         ||||
Db    4 CTGC 1
```

```
RESULT 6
BB207438
LOCUS       BB207438      731 bp    mRNA    linear    EST 31-AUG-2001
DEFINITION  BB207438 RIKEN full-length enriched, 0 day neonate thymus Mus
            musculus cDNA clone A430081P11 3' similar to L03547 Mouse Ikaros
            DNA binding protein (Ikaros) mRNA, mRNA sequence.
ACCESSION   BB207438
VERSION     BB207438.2  GI:15409900
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 731)
```

```
BASE COUNT       197 a    192 c    190 g    149 t
ORIGIN                           modified pbluescript KS(+) after bulk excision from Lambda
                                 FLC I."

        Query Match         23.3%;   Score 362;  DB 9;    Length 731;
        Best Local Similarity 84.3%;  Pred. No. 4.7e-65;                  3 others
        Matches 420;  Conservative  0;  Mismatches  75;  Indels   3;  Gaps   1;

Qy   417 CACTGGAGAACGGCCCTCCAGTGCAATCAGTGCGGGCCTCATTCACCCAGAAGGCAA 476
         |||||||||||||||||||||||||||||| |||||||| || ||||||||||||||
Db   234 CATGGGTGAACGGCCTTTCCAGTGCAACCAGTGTGGGGCCTCCTTTACCCAGAAAGGCAA 293

Qy   477 CCTGCTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAA 536
         ||| |||||||||||||| |||||||||||||||||| || ||||||||||||| ||||
Db   294 CCTCCTGCGGCACATCAAGCTGCACTCGGGTGAGAAGCCCTTCAAATGCCATCTTTGCAA 353

Qy   537 CTACGCCTGCCGCCGGAGGACGGCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAA 596
         ||||||||||||||||||||||||||||||| ||||| || |||||||| |||||||||||
Db   354 CTATGCCTGCCGCCGGAGGGACGGCCCTCACCGGCCACCTGAGGACGCACTCCGTTGGTAA 413

Qy   597 ACCTCACAAATGTGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGGAGGAACA 656
         ||||||||||||||||||||||||||| |||||| |||||||||||||| ||||||||||
Db   414 GCCTCACAAATGTGATATTGTGGCCGGAGCTATAAACAGCGAAGGTCTTTAGGAGGAGCA 473

Qy   657 TAAAGAGGCGCTGCCACAACTACTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCAGT 716
         ||||| ||||||||||||||||||||||||| |||||||||||||||||   ||||||||
Db   474 TAAAGAGGCGATGCCACAACTACTTGGAAAGCATGGGCCTTCCGGGCCA---TGTACCCAGT 530

Qy   717 CATTAAAGAGAAACTAAGCACAGTGAAATGGCAGAAGACCTGTCGAAGATAGGATCAGA 776
         |||||||||||||||||| |||| ||||||||||||||||||||||| |||||||||||
Db   531 CATTAAGGAGAAACTAACCACAACACAGAGATGGCAGAAGACCTGTCGAAGATAGGAGCAGA 590

Qy   777 GAGATCTCGTCCGTCGGACAGCTAGCAGTAATGTCGCCAAACGTAAGAGCTTCTATGCC 836
         ||| || || |||||||||||||||||| |||||||||||| |||||||||||||||||
Db   591 GAGGTCCCTTGTCCTGGACAGGCTGGCAAGCAATGTCGCCAAACGTAAGAGCTTCTATGCC 650

Qy   837 TCAGAAATTTCTTGGGACAAGGGCCTGTCCGACACGCCCTACGACAGTGCCACGTACGA 896
         ||||||||||||| ||| ||| || ||| |||||| || || || |||||||||||||
Db   651 TCAGAAATTTCTTGGTAAGGTTAAATGTTTGCGTCTTCTTAACCTAGAATTATGTGGGT 710

Qy   897 GAAGGAGAACGAAATGAT 914
         ||||||| |||| ||
Db   711 GTTCTAGACAAAAATTAT 728
```

```
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            ,Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
```

```
TITLE       RIKEN Mouse ESTs (Arakawa,T. et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8072391.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S. Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res . 11 (2), 281-289 (2001)
            Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
            Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..731
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A43008lP11"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     thymus"
                     /tissue_type="thymus"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCGAGTTAATTAAATTAAATTAATCCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
```

```
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new
  genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
  RIKEN integrated sequence analysis (RISA) system--384-format
  sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
  Computer-based methods for the mouse full-length cDNA
  encyclopedia: real-time sequence clustering for construction of a
  nonredundant cDNA library. Genome Res. .11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K. Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
  Computational Analysis of Full-Length Mouse cDNAs Compared with
  Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
  e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..669
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9930017A02"
                     /clone_lib="RIKEN full-length enriched, adult female
                     vagina"
                     /sex="female"
                     /tissue_type="vagina"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTGAGTTAATTAARTAATCCCCCCCCCCCCCC
                     3'].cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT    167 a    177 c    206 g    118 t    1 others
ORIGIN

Query Match         22.8%;
Best Local Similarity   91.4%;    Pred. No. 2.5e-63;
Matches 374;   Conservative 0;   Mismatches 35;   Indels 0;   Gaps 0;

Qy     1  ATGGATGCTGACGAGGTCAAGACATGCTTTCTCATCAGGGAAGGAAAGCCCCCCTGTA  60
          |||||| |||| ||| ||||||||||| |  || ||||||||| ||| ||||||  ||
Db   261  ATGGATGTCGATGAGGGTCAAGACATGTCCCAAGTTCAGGAAAGGAGAGCCCCCAGTC  320
Qy    61  AGCGATACTCCAGATGAGGGCGATGAGCCCGATACCCCGAGGACCTCCACCACC  120
          || || |||||||||| || |||||||| |  |||||||||| ||||| || |||
Db    61  AGCGATACTCCAGATGAAGGGGATGAGCCGATGAGCCCCATGTGTCCCGAGGACCTGTCCACCACC  120
Qy   321  AGTGACACTCCAGATGAAGGGGATGAGCCCATGAGCCCCATGTGTCCCGAGGACCTGTCCACTACC  380
          || || |||||||||| ||||||||||||| || |||| ||| ||||||||||||
Db   321  AGTGACACTCCAGATGAAGGGGATGAGCCCATGAGCCCCATGTGTCCCGAGGACCTGTCCACTACC  380

Qy   121  TCGGGAGGACAGCAGCAAAGTTCCAAGAGTCGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAG  180
          ||| ||| ||||||| |  |||||||||||  || || |||||||||||||||||||||||||
Db   381  TCTGGAGCCACAGCAGAACTCCAAGAGTGATCGAGGCCATGGCCAGTAATGTTAAAGTAGAG  440
Qy   181  ACTCAGAGTGATGAGAGAAGAATGGCGTGCCGTGCCTGTGAAATAATGGGGAGAAATGTGCCGAG  240
          |||||||||||||| ||| ||||||||||||||||||||||||||| |||||||||||||||
Db   441  ACTCAGAGTGATGAAGAGAATGGCGTGCCGTGCCTGTGAAATGATGGGGAGAAATGTGCAGAG  500
Qy   241  GATTTACGAAGTCTTGATCGTCCGGAGAGAGAAAATGAATGGCTCCCACAGGGACCAAGGC  300
          ||||||||||||| |||| || |||||||||||||||||||||||||||||||||||||
Db   501  GATTTACGAAGTCTGATGCTCCGGAGAGAGAAAATGAATGGCTCCCACAGGGACCAAGGC  560
Qy   301  AGTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTAAAGTGTGAT  360
          ||||||||||||| ||||||||||||||||||||||||||||||| |||||||||||||
Db   561  AGTCGGCTTTGTCAGGAGTTGGAGGCATTCGACTTCCTAACGGANAACTAAAGTGTGAT  620
Qy   361  ATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATCGTTCACAAAA  409
          ||||||||||||| |||||||||||||||||||||||||||||||
Db   621  ATCTGTGGGATCGTTTGCATCGGGCCCAATGTGCTCATCGTTCACAAAA  669


RESULT  8
AA866707/c
LOCUS       AA866707         936 bp   mRNA   linear   EST 16-MAR-1998
DEFINITION  vx85a10.r1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:1281978 5' similar to gb:L03547 Mouse Ikaros DNA binding
            protein (MOUSE);, mRNA sequence.
ACCESSION   AA866707
VERSION     AA866707.1  GI:2962152
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 936)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
  COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:673778
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 499.
FEATURES             Location/Qualifiers
     source          1..936
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1281978"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCGAAGTGGGAGCCGCGGTTTTTTTTTTTTTTTTT
                     3'], double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
```

COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 457   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 295.

FEATURES   Location/Qualifiers
     source   1..339
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1307771"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7D3-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).   cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGACGGCGGCCGCCCATTTTTTTTTTTTTTTTT-3'
                     ].   Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   69 a   102 c   75 g   93 t
ORIGIN

Query Match   21.0%;
Best Local Similarity   98.2%;   Score 325.4; DB 9; Pred. No. 1.5e-57; Length 339;
Matches 329; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   162 CAGTAATGTTAAAGTAGAGACTCAGAGTGATGAAGAGAATGGCGTGCCTGTGAAATGAA 221
          || || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   339 CACGAGGGTTAAAGTAGAGACTCAGAGTGATGAAGAGAATGGCGTGCCTGTGAAATGAA 280

Qy   222 TGGGGAAGAATGTGCCGGAGGATTTACGAATGCTTACGAGTTCTCGTC 281
          ||||||||||||||||||||||||||||||||||||||||||||| ||
Db   279 TGGGGAAGAATGTGCCGGAGGATTTACGAATGCTTACGAGTTCTCATC 220

Qy   282 CTCCCACAGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAA 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   219 CTCCCACAGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACCTTCTAA 160

Qy   342 CGGAAAACTAAAGTGTGATATCGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGT 401
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   159 CGGAAAACTAAAGTGTGATATCGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGT 100

Qy   402 TCACAAAGAGAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCTCATT 461
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 TCACAAAGAGAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCTCATT 40

Qy   462 CACCCAGAAGGGCAACCTGCTCCGGCACCATCAAGC 496
          ||||||||||||||||||||||||||||||||||||
Db   39 CACCCAGAAGGGCAACCTGCTCCGGCACCATCAAGC 5

RESULT 10
AA814418/c
LOCUS   AA814418   369 bp   mRNA   linear   EST 05-MAR-1998
DEFINITION nz12d08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1287567 3'
            similar to TR:O00598 O00598 HIK1. ;, mRNA sequence.

---

                         provided by Dr. Bertrand Jordan. Library went through two
                         rounds of normalization, and was constructed by Bento
                         Soares and M.Fatima Bonaldo."

BASE COUNT   212 a   241 c   267 g   216 t
ORIGIN

Query Match   22.5%;
Best Local Similarity   77.8%;   Score 348.8; DB 9;   Length 936;
Matches 498; Conservative   0; Mismatches 132; Indels 10; Gaps 6;

Qy   920 CCCACGTGATGGACCAAGCCATCAACACGCCATCAACTACCTGGGGC-CGAGTCCCTG 978
          || ||| ||| | |||||||||| ||| ||| |||||||| |||||||||
Db   916 CCACCGTAATGAACAGCAGGCCATTCAACAATTCCATAAACTACCTGGGGCTTGAGTCCCTC 857

Qy   979 CGCCGGCGGTGGTGCAGACGCCCCGGGGGGTTCCGAGTGGTCCCGGTCATCAGC--CCGA 1036
          || || || | | ||||| |||| ||| ||| | ||||||||| |||
Db   856 GCCCATTTGTGCGACACCCCCCCCGGTTAGCTACGAAGTGTTGCCAGTCATCAGCCTCCCA 797

Qy   1037 TGTACCAGCTGCACAGGCGGCTCGGAGGGCACCCGGCGCTCCAACCACTCGGCCCAGGACA 1096
          ||||||||||||||| | | | | | || || |||| ||| | | | ||||
Db   796 TGTACCAGCTGCACAAGCCCCCCAAGATGGCCCCCACGCTTCCAACCATCAAGAACA 737

Qy   1097 G--CGCCTGGGAGTACCTGCGTCTCTCCAAGGCCAAGTTGGTGCCCTCCGGAGCCCGA 1154
          | | ||||||| | || || | |||||| | ||| ||| | ||||| |
Db   736 GGACGCCGGTGATAACTTGCTGCTGCGTCCAAGGCCAAGTCGTGTCATCGGAGCCGAGA 677

Qy   1155 GGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGCAACACCGAGAGCAACACAGAGGAGCA 1214
          ||| ||||||||| ||||||||||||| ||| || | ||| || | | || ||
Db   676 GCCTTCCCCGAGCAACAGCTGCCAAGACTCCACAGATACAGAGAGCAAC-GCGGAGAACA 618

Qy   1215 GGGCAGGGTCTTATCTACCTGACCAACCACATCGCCCGACCGCGC---AACGGCTGTC 1271
          |||| ||| ||| || |||| || |||||||||| | || || | | | |
Db   617 GCCAGCGGCCTTATCTACCT-ACCAACCATCAACCCGCATGACGCAATGGGCTGGC 559

Qy   1272 GCTCAAGGAGGAGCACCCGCGCTACGACCTGTCTGCGCGCCTCCGAGAACTCGCAGGA 1331
          ||||||||||||||||||||||||||| | | | || ||||||||||||||||||
Db   558 TCTCAAGGAGGAGCAGCCGCGCTACGAGGTGCTGAGGGGCGCCTCCGAGAACTCGCAGGA 499

Qy   1332 CCGGCTCCGCGTGGTCAGCACCAGCGGGGAGCAGCAGATGAAGGTGTACAAGTCGCAACACTG 1391
          || || || |||||| ||||||| ||| | || || || |||||| |||||||||||||
Db   498 TGCCTTCCGTGTCGTCAGCACCAGCGGCGAGCAGCAGTCGAGGGTGTACAAGTCGCAACACTG 439

Qy   1392 CCGGGTGCTCTTCCTGGATCACGTCATGTACACCACCATCCACAGCGCCTTCCG 1451
          ||| |||||||||||||||||||| ||||| | | ||| |||||||||||
Db   438 CCGGTGCTCTTCCTGGATCACGTCATGTATACCATTCACAGCGGCCATGGGTTTCG 379

Qy   1452 TGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCGCCAGGACCGGTACGAGTTCTCGTC 1511
          || | |||| |||| || |||| || || || || |||||| | | | ||||||| |
Db   378 GGATCCCTTTGAGTGTAACATGTGTGGTTAATCACCACCAGGCAGTTACGAGTTCTCATC 319

Qy   1512 GCACATAACGGCGAGGGGAGCACCGCTTCCACATGAGCTAA 1551
          | ||||| ||| |||||||||||| || | || | || |
Db   318 CCATATCACGCGGGGGGAGCATCGTTACCACCTGAGCTAA 279

RESULT 9
AA808413/c
LOCUS   AA808413   339 bp   mRNA   linear   EST 19-FEB-1998
DEFINITION oa43h06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1307771 3'
            similar to TR:O00598 O00598 HIK1. ;, mRNA sequence.
ACCESSION   AA808413
VERSION   AA808413.1   GI:2877819
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 339)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)

```
ACCESSION   AA814418
VERSION     AA814418.1  GI:2884014
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 433   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1287567"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT    75 a     87 c     84 g    122 t      1 others
ORIGIN

Query Match             17.5%;  Score 271.4;  DB 9;  Length 369;
  Best Local Similarity   95.7%;  Pred. No. 2.4e-46;
  Matches  289;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy      208 GCCTGTGAAATGAATGGGAAGAATGCTGCGGAGGATTTACGAATGCTTAGTGCCTGCGGGA 267
            ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      369 GCCTGTGAAATGAATGGGAAGAATGCTGCGGAGGATTTACGAATGCTTGATGCCTGCGGGA 310

Qy      268 GAGAAAATGAATGGCTGTCCCACAGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGC 327
            ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      309 GAGAAAATGAATGGCNTCCACAGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGC 250

Qy      328 ATTCGACTTCCTAACGGAAAACTAAAGTGTGATATCTGTGGGATCATTGCATCGGGCCC 387
            |||||| ||||||||||||||||||||||||||||||||||  |||||||||||||||
Db      249 ATTCGACTCTCTAACGGAAAACTAAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCC 190

Qy      388 AATGTGCTCATGGTTCACAAAAGAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAG 447
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      189 AATGTGCTCATGGTTCACAAAAGAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAG 130

Qy      448 TGCGGGGCCTCATTCACCCAGGAAGGCAACCTGCTCCGGCACATCAAGTGCATTCCGGG 507
            ||| |||||||||||||||||||||||||||||||||||||||||||    |||
Db      129 TGC-GGGCCTCATTCACCCAGGAAGGCAACCTGCTCCGGCACATCAAGCCCTTCTCCCTG 71
```

```
Qy       508 GA 509
             |
Db        70 CA 69

RESULT 11
LOCUS       AW976805                 293 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  EST388914 MAGE resequences, MAGO Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW976805
VERSION     AW976805.1  GI:8168040
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 293)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 366
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1..293
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGO"
                     /note="Vector: pBluescriptSKm"
BASE COUNT    60 a     79 c     79 g     75 t
ORIGIN

Query Match             16.2%;  Score 251.6;  DB 9;  Length 293;
  Best Local Similarity   96.4%;  Pred. No. 2.9e-42;
  Matches  268;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy      233 GTGCGGAGGATTTACGAATGCTTGATGCCTC-GGGAGAGAAAATGAATGGCTCCACAGG 291
            |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db      293 GTGCGGAGGATTTACGAATGCTTGATGCCTCGGGGAGAGAAAATGAATGGCTCCACAGG 234

Qy      292 GACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTA 351
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      233 GACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTA 174

Qy      352 AAGTGTGATATCTGTGGGATCATTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGA 411
            ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      173 AAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGA 114

Qy      412 AGCCACACTGGAGAACGGCCCTTCCAGTCAATCAGTGCGGGGCCTCATTCACCCAGAAG 471
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      113 AGCCACACTGGAGAACGGCCCTTCCAGTCAATCAGTGCGGGGCCTCATTCACCCAGAAG 54

Qy      472 GGCAACCTGCTCCGGCACATCAAGTGCATTCCGGGA 509
            |||||||| |||||||||||||  |||   |
Db       53 GGCAACCTGCTCCGGCACATCAAGCCCTTCTCCCTGCA 16

RESULT 12
LOCUS       BM458447                1093 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6414077 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:547068
            5', mRNA sequence.
ACCESSION   BM458447
```

```
VERSION     BM458447.1  GI:18507487
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1093)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12126 row: j  column: 13
            High quality sequence stop: 663.
FEATURES    Location/Qualifiers
source      1..1093
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5497068"
            /clone_lib="NIH_MGC_85"
            /tissue_type="lymphoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.867 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT      345 a    242 c    274 g    230 t    2 others
ORIGIN

Query Match          15.4%;   Score 239.2;   DB 10;   Length 1093;
Best Local Similarity   66.5%;   Pred. No. 1.5e-39;
Matches  448;  Conservative   0;  Mismatches 190;  Indels 36;  Gaps  6;

Qy   338 CTAACGGAAAACTAAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCA 397
Db   380 CCAGTGGAAAGATGAACTGCGATGTGTTGGATTATCCTGCATCAGCTTCAATGTCTTAA 439

Qy   398 TGGTTCACAAAAGAAGCCACACTCGAGAAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCT 457
Db   440 TGGTTCATAAGCGAAGCCATACTGGTGAACGCCCATTCCAGTGTAATCAGTGTGGGGCAT 499

Qy   458 CATTCACCCAGAGAGGGCAACCTGCTCCGGCACACTCAAGCTGCATTCCGGGGAGAAGCCT 517
Db   500 CTTTACTCAGAAAGGTAACCTCCTCCGCCACATTAAACTGCACAGGGGAAAAACCTT 559

Qy   518 TCAAATGCCACCTCTGCAACTACGCCCTGCCGCCGGAGGACGGCCCTCACTGCCACCTGA 577
Db   560 TTAAGTGTCACCTCTGCAACTATGCATGCCAAAGAGAGATGCGCTCACGGGCCATCTTA 619

Qy   578 GGACGGCCACTCCGTTGGTAAACCTCACAAATGTGGATATTGTGCCGAAGCTATAAACAGC 637
Db   620 GGACACATTCTGTGGAGAAACCCTACAAATGTGAGTTTTGTGGAAGGAGTTACAAGCCAGA 679

Qy   638 GAACGTCTTTAGAGGAACATAAAGAGCCTGCCACAACTACTTGGAAAGCATGGGCCTTC 697
Db   680 GAAGTTCCCTTGAGGAGGACGACAAGGAGGCGGCCTGCCGTACATTTCTTCGAGCACTGACC--C 737

Qy   698 CGGGCACACTGTACCCAGTCATTAAAGGAAGAAGAAACTAAGCACAGTGAAAATGGCAGAAGACC 757
Db   738 AGGGGACACTGCAAGTGCGGAGGCGAAGCAAGACACATCAAAGCA-------------------- 777

Qy   758 TGTGCAAGATAGGATCAGAGAGATCTGCCGTGCTGGACGACTAGCAAGTAATGTGCCCCA 817
Db   778 -----GAGATGGGAAGTGAAAGAGCTTCGTACTGGACCGACAGATTAGCAAGCAATGTGGCAA 832
```

```
Qy   818 AACGTAAGAGCTCTATGCCTCAGAAATTCTTGGGACAAGGGCCTGTCCGACACGCCCT 877
Db   833 AACG--AAAAGCTCAATGCCTCAGAAATTCATTGGTGAGGGCGCCACTGCTTGATGTCA 891

Qy   878 ACGA-----CAGTGCCACGTACGACGAAGGAGGAACCAAAATGATGAGTCCCACGTGATGGA 932
Db   892 ACTATAATTCAGTTACATGTATGAGAAAA-AATGAGTCCATACAACGGACCCGCCATGATGGA 950

Qy   933 CCAAGCCATCAACAACACGCCATCAACT--ACCTGGGGCCGGCCAGTCCTGGCCCCGCCTGGTG 990
Db   951 CCAANCCATCAATAACGGCATCAGCTTATCCTCTGCGCCGCCCCAAGCCCCTGGCCCCCTTGGGG 1010

Qy   991 CAGACGCCCCGGG 1004
Db  1011 TCAAAAACGCCTGG 1024
```

RESULT 13
BF731127/c

```
LOCUS       BF731127               704 bp    mRNA    linear   EST 08-JAN-2001
DEFINITION  mab81ell.y1 NCI_CGAP_BC3 Mus musculus cDNA clone IMAGE:3976797 5'
            similar to SW:IKAR_MOUSE Q03267 DNA-BINDING PROTEIN IKAROS ;, mRNA
            sequence.
ACCESSION   BF731127
VERSION     BF731127.1  GI:12048991
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 704)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            MGI:1476829
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 445.
FEATURES    Location/Qualifiers
source      1..704
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:3976797"
            /clone_lib="NCI_CGAP_BC3"
            /tissue_type="marginal zone B-cell tumor"
            /lab_host="DH10B (T1-resistant)"
            /note="Organ: lymphocytes (flow-sorted); Vector:
            pCMV-SPORT6 (Life Technologies); mRNA made from
            flow-sorted lymphocytes, cDNA made by oligo-dT priming.
            Directionally cloned. Average insert size 1.8 kb. Primary
            library, non-amplified.   cDNA Library Preparation: David
            B. Krizman, Ph.D."
BASE COUNT      162 a    200 c    186 g    154 t    2 others
ORIGIN

Query Match          15.38;   Score 237.4;   DB 10;   Length 704;
Best Local Similarity   83.6%;   Pred. No. 3.2e-39;
Matches  281;  Conservative   0;  Mismatches  52;  Indels 3;  Gaps 1;

Qy  1219 AGCGGTCTTATCTACCTGACCAACCATCGCCCGACGCGGC--AACGGCGTGTCGCTC 1275
```

Query Match
Best Local Similarity   14.9%;   Score 231.8; DB 10; Length 571;
                                       81.18; Pred. No. 4.4e-38;
Matches 305; Conservative 0; Mismatches 68; Indels 3; Gaps 3;

```
Qy    1 ATGGATGCTGACGAGGTCAAGACAATGTCTTTCTCATCAGGGAAGGAAAGCCCCCCTGTA  60
        ||||||| || ||||||||||||||||||||||         |||| |||| ||||||||||
Db  185 ATGGATGTCGATGAGGGTCAAGACAATGTCCCAAGTTTCAGGAAAGGAGAGCCCCCAGTC  244

Qy   61 AGCGATACTCCAGATAGGGGCGATGAGCCCATGCCCGAGGACCTCTCCACCACC  120
        ||||| || || |||||| |||||| ||| |||| || || |||||||||||||
Db  245 AGTGACACTCCANATGAAGGGGATGAGCCCATGCCTGAGGACCTGTTCACTACC  304

Qy  121 TCGGGAGGACAGCAAAGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAG  180
        || |||| |||||||||||||||||||||||||  |||||||||| |||||||||||||||
Db  305 TCTGGAGCACAGGAGAACTCCAAGAGTGATCGAGCCAGTAATGTTGAAGTAGAG  364

Qy  181 ACTCAGATGATGAACAGAATGGGGCTGCCTGTGAAATGAAAATGTGCGGAG  240
        |||||||| ||||| |||||| |||| || ||||| ||| ||  || |||||
Db  365 ACTCAGATGATGAAGAGAATGGGGCCGCCCGTGATATGAATGGGGAAGAATTGTGCGAA  424

Qy  241 GATTTACGAAGTCTTGATGCCTCGGGAGAGAAATGAATGGT-CCACATGGATCCAGGG  300
        ||||||||||||| ||||| |||||||| ||||| |||||||| |||||||||||||||
Db  425 GATTTACCAATGCTTGATTGCCTCCGGAAAGAAAAAATAATGGT-CCACATGGATCCAGGG  483

Qy  301 AGCTCGGCTTGTCTGGGAGTTGGAGGCATTGGACTTCCTAACGGAAAACTAAAGTGTGAT  360
        ||| | ||||||| |||||||||||||||| ||| ||||||||||||| ||||||||||
Db  484 CGCTTCGGCTTGTCAGGAGTTGGA-GCATTGCGACTTCCTTACGGAAAACT-AAGTGTGAT  541

Qy  361 ATCTGTGGGATCATTT  376
        ||||||| |||||||
Db  542 TTCTGTGGATTCGTTT  557
```

RESULT 15
AI323339
LOCUS   AI323339   451 bp   mRNA   linear   EST 23-DEC-1998
DEFINITION mi06e03.x1 Soares mouse placenta 4NbMPl3.5 14.5 Mus musculus cDNA
           clone IMAGE:459676 3' similar to gb:L03547 Mouse Ikaros DNA binding
           protein (MOUSE);, mRNA sequence.
ACCESSION AI323339
VERSION   AI323339.1   GI:4057768
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 451)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
           WashU-HHMI Mouse EST Project
           Washington University School of MedicineP
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:276564
           This clone was previously sequenced on the 5' end only, this new
           data is from the 3' end
           High quality sequence stop: 429.
FEATURES   Location/Qualifiers
     source   1..451
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
```
Db  606 AGCGGCCGTATCTACCTAACCACCCACATCACCCGCCGCAATGCCACGCAGCGCTGGCTCTC  547
```

```
Qy 1276 AAGGAGGAGCACCGCCTACGACCTGCTGCCGCCGCCTCCGAGAACTCGCAGGACGCG  1335
        ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||| ||
Db  546 AAGGAGGAGCAGCGCCGCCTCGAGGTGCTGAGGCCGCCTCAGAGAACTCGCAGGACGCC  487

Qy 1336 CTCCGCGTGGTCAGCACCAGGGGGGGAGCAGAGTGAAGGTGTACAAGTGCGAACACTGCCGG  1395
        || ||||||||||||| || |||||||||||||||||||| ||||||| |||||||||||||||
Db  486 TCCCGTGTGGTCAGCACGAGTGCGGCAGCAGCTGAAGGTGTACAAGTGCGAACACTGCCGC  427

Qy 1396 GTGTCTTCCTGGATCACGTCATGTACACCATCCACATGGGGCTGCCACGGCTTCCGTGAT  1455
        ||||||||||||||| ||||||| || ||||||||| |||||||||||||||||| ||
Db  426 GTGTCTTCCTGGATCACGTCATCACGTATACCATTCACATGGGGCTTCCGGGAT  367

Qy 1456 CCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGTACGAGTTCTCGTGCAC  1515
        || |||| ||||||||||||||||| || || |||||||| ||||||| |||||||||
Db  366 CCCTTTGAGTGTAACATGTGTGGTTATCACAGCCAGGACAGGTACGAGTTTCTCATCCAT  307

Qy 1516 ATAACGCGAGGGGAGCACCGCCTCCACATGAGCTAA  1551
        |||||||||||||||||||| || ||||||| |||||
Db  306 ATCACGCGGGGGGAGCATCGTTACCACAGTGAGCTAA  271
```

RESULT 14
BG089790
LOCUS   BG089790   571 bp   mRNA   linear   EST 26-JAN-2001
DEFINITION mab8le11.x1 NCI_CGAP_BC3 Mus musculus cDNA clone IMAGE:3976797 3'
           similar to SW:IKAR_MOUSE Q03267 DNA-BINDING PROTEIN IKAROS ;, mRNA
           sequence.
ACCESSION BG089790
VERSION   BG089790.1   GI:12572353
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 571)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Other_ESTs: mab8le11.y1
           Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov

           MGI:1476829
           Possible reversed clone: similarity on wrong strand
           Seq primer: -40UP from Gibco
           High quality sequence stop: 244.
FEATURES   Location/Qualifiers
     source   1..571
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3976797"
                     /clone_lib="NCI_CGAP_BC3"
                     /tissue_type="marginal zone B-cell tumor"
                     /lab_host="DH10B (Ti-resistant)"
                     /note="Organ: lymphocytes (flow-sorted); Vector:
                     pCMV·SPORT6 (Life Technologies); mRNA made from
                     flow-sorted lymphocytes, cDNA made by oligo-dT priming.
                     Directionally cloned. Average insert size 1.8 kb. Primary
                     library, non-amplified.   cDNA Library Preparation: David
                     B. Krizman, Ph.D."
BASE COUNT   140 a   148 c   171 g   111 t   1 others
ORIGIN

```
                /db_xref="taxon:10090"
                /clone="IMAGE:459676"
                /clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
                /sex="unknown"
                /tissue_type="placenta"
                /dev_stage="adult"
                /lab_host="DH10B"
                /note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
                with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                1st strand cDNA was primed with a Not I - oligo(dT) primer
                [5'
                TGTTACCAATCTGAAGTGGGAGCCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTT
                T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. Library
                went through one round of normalization, and was
                constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      115 a      121 c       141 g        74 t
ORIGIN


Query Match       14.8%;    Score 229.8;  DB 9;   Length 451;
Best Local Similarity  88.6%;   Pred. No. 1.1e-37;
Matches  249;   Conservative   0;  Mismatches  32;   Indels   0;   Gaps   0;

Qy    1 ATGGATGCTGACGAGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAGCCCCCTGTA  60
          |||||| || ||||||||||||||||||       |||||  |||| ||||||||||
Db  171 ATGGATGTCGATGAGGTCAAGACATGTCCCAAGTTTCAGGAAAGGAGAGCCCCCCAGTC 230

Qy   61 AGCGGATACTCCAGATGAGGGCGATGAGCCCATGCCCATCCCCGAGGACCTCTCCACCACC 120
          |||| |||||||||| || |||||||||  ||||||||||  |||| |||||||| ||
Db  231 AGTGACACTCCAGATGAAGGGGATGAGCCCATGCCCTGTCCCTGAGGACCTGTCCACTACC 290

Qy  121 TCGGGAGGACAGCAAAGCTCCAAGATGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAG 180
          || |||| ||||| ||||||||||||  |||  | | ||||||| |||||||||||||
Db  291 TCTGGAGCACAGCAGAACTCCAAGAGTGATCGAGGCCATGGCCAGTAATGTTAAAGTAGAG 350

Qy  181 ACTCAGAGTGATGAAGAAGAATGGGCGTGCCTGTGAAAATGAATGGGGAAGAATGTGCGGAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  351 ACTCAGAGTGATGAAGAAGAATGGGCGTGCCTGTGAAAATGAATGGGGAAGAATGTGCAGAG 410

Qy  241 GATTTACGAATGCTTGATGCCTCGGGAGAGAAAATGAATGG 281
          |||||||||||||||||||||||||||||||||||||||||
Db  411 GATTTACGAATGCTTGATGCCTCGGGAGAGAAAATGAATGG 451


Search completed: August 28, 2002, 09:02:15
Job time: 6895 sec
```